

**A Modeling Approach toward Developing
Small Area Demographic Projections**
(draft 100612)

By

Simon Choi
Southern California Association of Governments
818 West 7th Street 12th Floor
Los Angeles, CA 90017
* choi@scag.ca.gov
* 213-236-1849

Paper presented at the 2012 Southern Demographic Association Annual Meeting,
October 10-12, 2012, Williamsburg, VA

Abstract

Projections of small area population size play an important role in determining the diverse future needs of local communities in the regional plan framework. With availability of the more detailed projections of demographic and housing growth at the small areas, local and regional planners would develop more relevant local and regional plans. The population and demographic change of small areas is mostly driven by both the regional economic-demographic influence and small area demographic processes (births, deaths, and migration) in a spatio-temporal context. The popular cohort-component method is not easily applicable to population projections of small area (e.g., census tracts, transportation analysis zones) for a couple of reasons: (1) the historical and current trends in vital statistics and migration of the small areas are not easily available; (2) it is difficult to independently develop reasonable migration assumptions of small areas. A proposed approach is to derive population projections of local communities through the enhanced linkage of small area housing growth and regional demographic processes. A proposed small area modeling approach is as follows: (1) project the regional population growth using the cohort-component model; (2) project the small area housing units using diverse models (e.g., trend projections, simple land use models, complex land use models, envision scenario); (3) convert the small area housing growth into population growth using the housing unit method; (4) disaggregate the small area population into their demographic characteristics (e.g., age, race/ethnicity). This study discusses key components of the proposed modeling approach and suggests that the proposed modeling approach can be a useful scenario testing tool for urban and regional planning.

Keywords: population, demographic projections, small area, growth scenarios, spatial distribution of age and race/ethnicity.

Introduction

The small area population projections are important in understanding the diverse community service needs of the future (Klosterman, 1990). Although the population size of the small area might be a useful indicator for measuring the community needs, the detailed demographic projections, if available, would be able to more accurately estimate the community service needs. This study presents a coherent modeling framework for projecting the population size and demographic characteristics of the small area within the metropolitan planning context.

The population and demographic change of small areas is mostly driven by both the regional economic-demographic influence and small area demographic processes (births, deaths, and migration) in a spatio-temporal context. The small area demographic process, in particular, small area migration, is volatile along the national and regional economic cycles and hard to forecast as a major component of small area population projections. The expected small area migration would be easily absorbed into small area population projections with the expected small area demographic processes assuming that the timely housing supply is available for the projected in-migrants. If the timely housing supply is not readily available and the small area migration projection remains unchanged, local communities would experience a lot of changes in a small area population-housing relationship (e.g., higher household size) and a lower housing quality (e.g., household overcrowding). It is important for the small area demographic projections to consider both the demographic process at the large area level and the availability of housing at the small area level as part of a small area demographic modeling framework.

In fact, the previous study emphasized the importance of population or housing at the different levels of geography (Field and MacGregor, 1987). The study indicated that housing become a key measure for smaller levels of geography, while population for a larger levels of geography. The reason might be related to lack of necessary data and reliable assumptions. For example, the popular cohort-component method is not easily applicable to population projections of small area (e.g., city or census tracts) for a couple of reasons: (1) the historical and current trends in vital statistics and migration of the small areas are not easily available; (2) it is difficult to independently develop reasonable migration assumptions of small areas. An alternative approach is to derive population projections of local communities through the enhanced linkage of small area housing growth and regional demographic processes.

Small area housing growth can be easily used to determine migration and population projections of local communities from the local planning perspective. Local planners cannot easily access the small area demographic data due to its limited availability, but are familiar with housing development permits and process, the existing local general plan, zoning codes, and other land use regulations. They, at local jurisdictions, are monitoring the land use changes and housing development on a daily basis. They are charged with envisioning the future housing growth of the community. If there is opportunity, they tend to translate the future housing development into population and demographic projections, instead of vice versa.

This study presents a modeling approach toward developing the long term projection of total population and the key demographic characteristics (e.g., age, race/ethnicity) at the transportation analysis zone level. A proposed small area modeling approach is as follows: (1) project the regional employment, population, and household growth; (2) allocate the regional housing growth into the small area; (3) convert the small area housing growth into population growth using; (4) disaggregate the small area population into their demographic characteristics. This study focuses on the fourth stage of projecting the small area demographic characteristics and presents the multi-nomial logit regression method to project the small area demographic characteristics utilizing the past trend of those demographic components of population at the small area.

Overview

The cohort-component model is a widely used population projection model in the world due to its capability to produce key demographic characteristics including age, gender, and race/ethnicity (Smith et al, 2001). US Census Bureau has used the cohort-component method to produce the national population projections. The forecasting accuracy was relatively high. The cohort-component focused on age and gender at the early period of application, later added one more dimension of race/ethnicity with its increasing importance. There is a recent effort of adding more detailed demographic characteristics of immigrant residents (Myers & Pitkin, 2011). The additional characteristics include the birth place, length of stay in the US, of the US immigrants.

While the cohort-component model produces useful information of demographic characteristics of the projected population, it tends to target the larger levels of geography. The minimum level of geography is usually the county in the United States due to the availability of the necessary information of births, deaths, and migration. In particular, migration data is not readily available and might not track the gross migration (e.g., inflows and outflows of migration) in many cases. There is a wide range of studies of how to develop migration assumptions (Plane and Rogerson, 1994). The studies try to figure out the major determinants of migration at both the large area (usually county or state) level or above. The most popular determinant of migration would be job growth. Lowry used the job to population linkage to develop metropolitan growth model (Lowry, 1964). The job-population (or migration) linkage is still a dominant modeling framework in the urban and land use modeling field.

There is a growing demand of more detailed demographic characteristics at the very small area level in the field of business demography and transportation demand modeling and business demography. The private vendors in the economic and demographic field tend to produce short and long term population and demographic projections for clients of public and private sectors. These projections tend to reflect the business and public facility demand due to inclusion of population related variables, and are used to identify the optimal business and public facility location. The methodology and assumptions underlying the small area demographic projections are not well known to the general public due to its nature of business (Smith et al, 2001). The demographic characteristics of projected population at the small area level are directly related to diverse human

behavior including travel (trip generation, trip distribution, trip assignment, and mode choice), household formation, homeownership, health care, retirement, etc. With the reasonable estimates of the demographic characteristics, we might easily derive the implication for housing, travel, health care, retirement, etc. Because of the strong linkage of demographic characteristics to the diverse human behavioral activities, a regional planning organization tend to use that linkage to understand the future community needs and to derive policy options to accommodate the future service needs.

In contrast to the large area employment-population (migration) dynamics, the small area's demographic characteristics might not be easily figured out without knowing the overall population and housing growth of the small areas. A good example is a housing unit method. The housing unit method is a conceptually clear and theoretically sound population estimation method in the U.S. (Smith and Lewis, 1982). Since there is no officially mandated population registration system in the U.S., the population estimate is oftentimes derived using the housing growth. Housing estimate is relatively easy to collect through the local government. According to the housing unit method, the housing growth is translated to population through the necessary conversion process. There is a risk of the volatile conversion factors, such as housing occupancy rate, household size, or the share of group quarter population to total population, but the housing unit method is widely used due to the easiness of the data accessibility and method application.

The housing unit method is sometimes further extended to project the small area population in the metropolitan planning process (Southern California Association of Governments, 2008). As part of the baseline population projection development for the city level or below, the trend extrapolated housing growth is translated into population through the conversion process. A few conversion factors might not be stable during the projection period, and they are subject to major review. The housing unit method is well taken by local planners. Local planners cannot easily access the small area demographic data due to its limited availability, but are familiar with housing development permits and process, the existing local general plan, zoning codes, and other land use regulations. They, at local jurisdictions, are monitoring the land use changes and housing development on a daily basis. They are even charged with envisioning the future housing growth of the community. If there is opportunity, they tend to translate the future housing development into population and demographic projections, instead of vice versa.

The demographic characteristics of the small area may be projected using three different approaches. The first approach is the iterative proportional fitting (IPF) approach, and might be one of the most popular approaches currently used in the world. The IPF technique was first introduced by Deming and Stephan in 1940 and proved by the rigorous research of Fienberg (1970). IPF approach is preferred due to its computational speed, numerical stability and algebraic simplicity (http://en.wikipedia.org/wiki/Iterative_proportional_fitting). The IPF approach is also well recognized in the field of small area demography (Kanaroglou et al, 2009; Rees et al, 2004; Simpson and Tranmer, 2005). For example, the demographic characteristics of the small area could be determined using the reference region's population and demographic projections, and the small area's population projections. The only missing element is the

detailed demographic characteristics, probably age and racial/ethnic composition of the projections populations at the small area. The IPF uses the base year's age and racial/ethnic composition of the small area population as reference for the future age and racial/ethnic composition of the small area population. The IPF approach would be every effective in developing the relatively short term projections of the demographic characteristics due to the temporal continuity of the demographic characteristics of the "mature" small area, while it might be limited in projecting the long term projections for the "emerging" small area..

The second approach is a modeling approach. The modeling approach is often observed from the typical land use modeling process. The land use model tends to assign specified households with selected demographic characteristics (e.g., household income) to the small areas using residential location models (Pagliara, Preston, and Simmonds (eds), 2010; Brail and Klosterman (eds), 2001). There are good examples. The DRAM-EMPAL was developed by Putman in 1971, and was widely applied to the metropolitan land use modeling in 1980s and 1990s (Putman, 2010). The households by different income category are allocated to small areas using the small area zone's attractiveness and transportation accessibility. Through this kind of modeling practice, households of different income category are projected as a result of land use and/or transportation investment policy options. The modeling tradition has been carried over to the newly available land use modeling practice (e.g., PECAS, UrbanSim, DELTA, MUSSA II,). The new modeling approach tends to assign households with more demographic characteristics to small area zones according to policy alternatives. For example, Urbansim produces the small area population size and the demographic characteristics of households by income, age of head, household size, presence of children, and housing type. PECAS also produces key demographic characteristics of households, including household income, household size, status of households as senior households (whether the household are composed of population of 65 years old or more). The modeling approach does not produce a comprehensive dataset, but a limited number of key demographic characteristics related to households. In order to produce additional demographic variables of population, the statistical approach would be needed.

The third approach is a statistical approach (Cho, 2006; Kanaroglou et al, 2009; Eluru et al, 2008). The typical modeling process is to allocate the projected large area (e.g., county or metropolitan area) population into the small area (e.g., census tract, or census block group). The commonly applicable method of developing the demographic components of the small area population in the top-down statistical approach is the multinomial logit regression method. The regression coefficients might be derived using the individual data set (Eluru et al, 2008) or aggregate zonal database (Cho, 2006; Kanaroglou et al, 2009). The historical databases are used to extrapolate the historical trend of the small area demographic characteristics (Cho, 2006). As a result of the top-down approach, the small area database becomes consistent with the large area's demographic pattern, and the small area population size, and also the historical pattern of the demographic characteristics. Although this top-down approach presents major strength in producing the consistent small area dataset, it would not reflect the comparable demographic characteristics of a certain area to be developed. For example, a

transit-oriented development around the transportation station might result in the new residential and commercial development, and gentrification of replacing existing low income or ethnically minority residents with new middle income and professional job residents. The locally unique development related demographic changes are not properly reflected in the database. The alternative method in the top-down statistical approach is the locally weighted regression (LOESS), or LOWESS (locally weighted scatterplot smoothing). LOESS, originally proposed by Cleveland (1979) and further developed by Cleveland and Devlin (1988). LOESS combines much of the simplicity of linear least squares regression with the flexibility of nonlinear regression. It does this by fitting simple models to localized subsets of the data to build up a function that describes the deterministic part of the variation in the data, point by point.

(<http://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm>)

Park (2009) produced the reasonable small area demographic characteristics (e.g., age and racial/ethnic composition) using LOESS. The model showed a high level of the goodness of fit by showing a high R-square, and the model results were successfully validated against the actual dataset. The initial application of this technique provides a potential for the future use and requires a further research.

Modeling Framework

The following is a proposed modeling approach toward developing the long term projection of the key demographic characteristics (e.g., age, race/ethnicity) at the transportation analysis zone level (see figure 1). First, project the regional employment, population, and household growth using the employment, demographic, and household projections models. Second, allocate the regional housing and employment growth into the small area through the simple trend extrapolation, residential location model, or the envisioning process. For example, the envisioning process considers urban growth scenarios (e.g., urban concentrated, suburban sprawl) and/or alternative smart growth techniques (e.g., transit-oriented development, mixed use development, employment center development, etc.). Third, convert the small area housing growth into population growth using the housing unit method. The conversion factors including the occupancy rate, household size, and the share of the group quarters population need to be fully analyzed. Fourth, disaggregate the small area population into their demographic characteristics using the multi-nomial logit regression method utilizing the past trend of those demographic components of population at the small area. The historical change in the pattern of age or ethnic compositions at the TAZ level plays a key role in determining the future patterns of TAZs. The preliminary model results are controlled to the region wide projection. At the end, we would be able to assess the spatial pattern of age and ethnic composition of the small area population using the dependency ratio or diversity index. We can use the projected age and racial/ethnic distribution for the environmental justice analysis as required by the regional transportation planning process. This study focuses on the fourth stage of projecting the small area demographic characteristics.

This study presents a modeling process for the age and racial/ethnic composition of projected population at the transportation analysis. The more hierarchical zones would be a better option for developing more refined dataset due to increased possibility of

incorporating the locally unique situation. A small area modeling approach in this study is based on two hierarchical zones (region and transportation analysis zone). This study focuses on how to disaggregate the small area population into their demographic characteristics. The proposed small area secondary variables allocation model (SASVAM) designed to produce secondary variables at TAZ contains four major features: (1) reflect the historical pattern of the target demographic characteristics, (2) control to TAZ population and County demographic characteristics, (3) maintain consistency among the projected demographic characteristics, (4) maintain the monotonous pattern of projected demographic characteristics. The key feature of the proposed small area forecasting model is to utilize the historical pattern of the target demographic characteristics of the populations, while controlling for the county pattern of demographic characteristics of population (Cho, 2006). The changed historical pattern of the target demographic characteristics is determined by aggregated individuals as probabilistic choice. For example, a small area with a rapid population aging in the past, then it tends to continue according to the projected county level aging pattern. This approach is similar to the synthetic technique in the demographic field, but has strength in reflecting the historical pattern of the small area as part of the modeling framework.

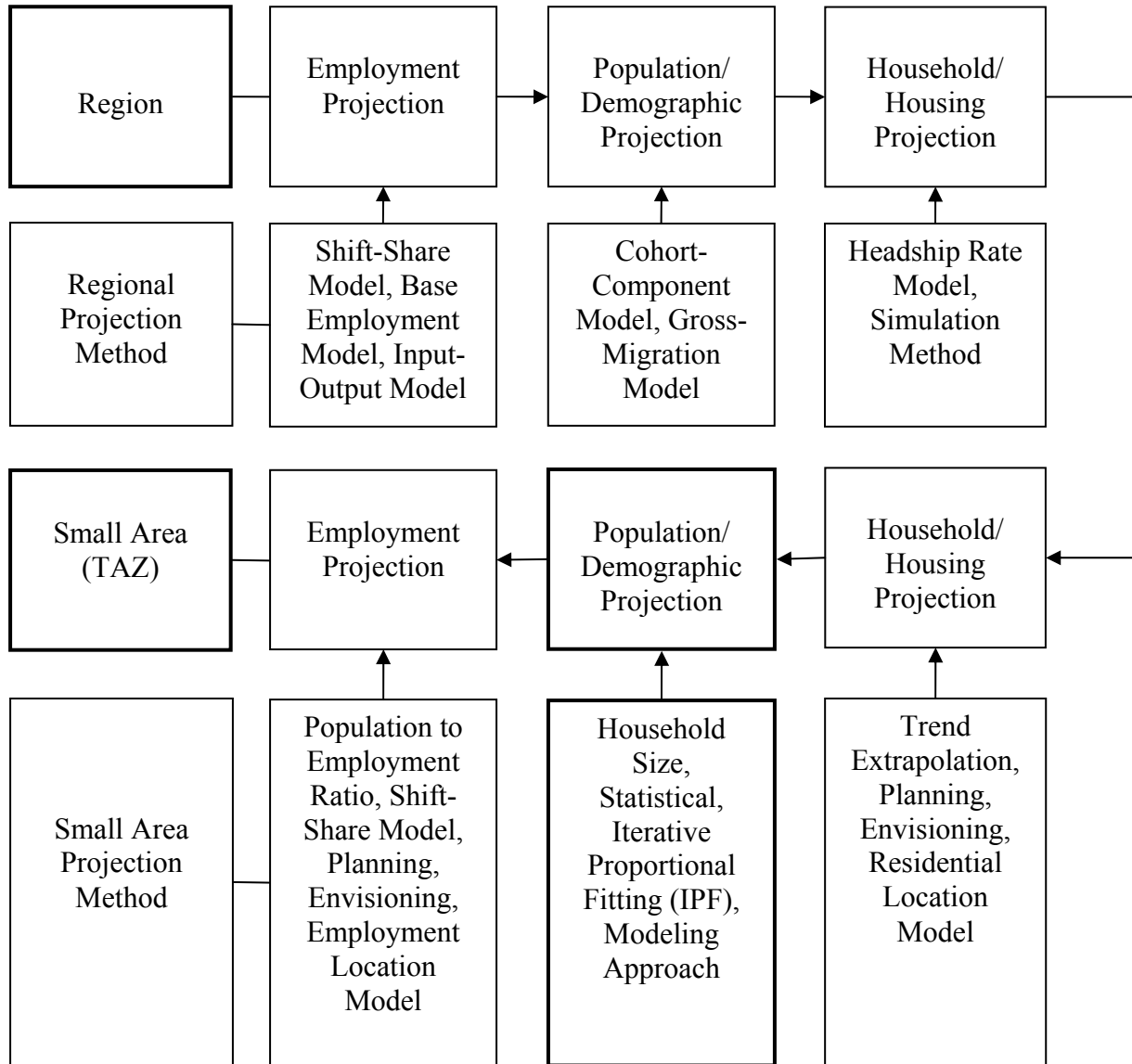


Figure 1. A Modeling Framework for the Regional and Small Area Demographic Projections

Data and Methods

The study area covers the whole Southern California Association of Governments (SCAG) region, comprised of six counties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, and Ventura. The SCAG region encompasses 191 cities, 38,000 square miles, and over 18 million people. SCAG is the largest metropolitan government organization (MPO) in the United States. SCAG is mandated by the federal and state governments to develop regional plans for transportation, growth management, housing development, air quality and other issues of regional significance.

The spatial unit of analysis in this study is primarily census tract (CT) for the statistical modeling and the transportation analysis zone (TAZ) for the model application. The current TAZ is developed using the 2000 Census Tract. Except for the outlying areas where the size of census tract is so large that it should be split into two or more TAZs. The size of TAZs is equivalent to that of census tracts. These TAZs are aggregated to 55 Regional Statistical Areas (RSAs) and 6 Counties. .

A multi-nomial logistic regression model is used to estimate the changing composition of the age and racial/ethnic groups at the census tract level in Southern California six counties between 1990 and 2000. A multi-nomial logistic regression describes the relationship between a categorical multi-nomial response variables and a set of predictor variables (Menard, 2002; Pampel, 2000; Liao, 1994). The model estimates the distribution probability of a certain age group or racial/ethnic group in 2000 as a function of age, racial/ethnic group, and other socioeconomic factors in 1990. The probability is transformed to a logit form so that there is a linear relationship between independent variables and the dependent variable. The logits of the unknown multi-nomial probabilities (i.e., the logarithms of the odds) are presented in the following form.

$$Prob(y = j) = \frac{e^{\sum_{k=1}^k \beta_{jk} x_k}}{1 + \sum_{j=1}^{J-1} e^{\sum_{k=1}^k \beta_{jk} x_k}}$$

Where

$j = 1, 2, 3, \dots, J-1$,

$Prob(y=j)$ = distribution probability of a certain age group or racial/ethnic group in 2000,

x_k = independent variables,

β_{jk} = estimated coefficients,

This study uses the 1990 and 2000 census data, Summary File 3 (SF3) and Census Transportation Planning Package (CTPP), which are collected for the census tract, taken from the sample, long-form questionnaires. 1990 census tract data was converted to 2000 census tract equivalent data using the land area method. First, the age composition of population in 2000 is determined by 1990 age composition of population and other related variables (see tables 1 and 2). The age group of population in 1990 and 2000 is categorized into seven (0-4, 5-15, 16, 17, 18-24, 25-64, 65+) as required by the transportation modeling process. The age groups of population in 1990 are processed as independent variables, while the age groups of population in 2000 are processed as dependent variables. The population of age 65+ in 2000 is treated as a reference dependent variable. Additional variables including 1990 Hispanic population, 1990 median household income, 1990 employment, 1990 population density per square mile are also added as independent variables due to its potentially significant influence on age composition of population in 2000. Second, the racial/ethnic composition of population in 2000 is determined by 1990 racial/ethnic composition of population in 1990 and other

related variables (see tables 1 and 2). The race/ethnicity of population in 1990 and 2000 is categorized into six (Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic American Indian, non-Hispanic Asian, non-Hispanic Others), while treating non-Hispanic Other population in 2000 as a reference dependent variable. Additional variables including 1990 population of age 25-64, 1990 median household income, 1990 employment, 1990 population density per square mile are also added as independent variables due to its potentially significant influence on the racial/ethnic composition of population in 2000.

Table 1. Description of Independent and Dependent Variables

Variable		Description
Dependent Variable		
1. Seven age groups in 2000 (a reference variable is an age group of 65 years or older) 2. Six racial/ethnic groups in 2000 (a reference variable is a Non-Hispanic Others category)		Probability of a person to belong to one of seven age groups or one of six racial/ethnic groups in 2000
Independent Variable		
Age in 1990	Age 0-4 Age 5-15 Age 16 Age 17 Age 18-24 Age 25-64 Age 65+	Population of Age0-4 Population of Age 5-15 Population of Age 16 Population of Age 17 Population of Age 18-24 Population of Age 25-64 Population of Age 65+
Race/ethnicity in 1990	Hispanic Non-Hispanic White Non-Hispanic Black Non-Hispanic American Indian Non-Hispanic Asian Non-Hispanic Others	Hispanic population Non-Hispanic White Population Non-Hispanic Black Population Non-Hispanic American Indian Population Non-Hispanic Asian Population Non-Hispanic Other Population
Income in 1990	Median Household Income	Median Household Income
Employment in 1990	Employment	Number of jobs by place of work
Population density in 1990	Population density	Population per square mile

Source: US Census Bureau, 1990 and 2000 Census Summary File 3 and 1990 CTPP.

Table 2. Descriptive Statistics of Independent and Dependent Variables

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Age 0_4_90	3,402	357	213	0	1,924
Age 5_15_90	3,402	675	391	0	4,041
Age 16_90	3,402	56	37	0	336
Age 17_90	3,402	59	38	0	430
Age 18_24_90	3,402	499	351	0	6,795
Age 25_64_90	3,402	2,236	999	0	10,176
Age 65_over_90	3,402	419	323	0	4,970
Median household income_90	3,402	51,014	22,474	6,472	194,191
Employment_90	3,402	2,012	4,055	3	78,655
Population density_90	3,402	9,029	9,087	0	87,449
Age 0_4_00	3,402	373	236	0	2,410
Age 5_15_00	3,402	872	533	0	5,919
Age 16_00	3,402	68	46	0	402
Age 17_00	3,402	70	48	0	429
Age 18_24_00	3,402	483	381	0	7,914
Age 25_64_00	3,402	2,509	1,176	0	12,194
Age 65_over_00	3,402	481	358	0	4,714
Hispanic_90	3,402	1,407	1,362	0	8,762
Non-Hispanic White_90	3,402	2,141	1,587	0	14,347
Non-Hispanic Black_90	3,402	345	717	0	6,829
Non-Hispanic Indian_90	3,402	19	30	0	1,164
Non-Hispanic Asian_90	3,402	381	468	0	4,468
Non-Hispanic Other_90	3,402	9	15	0	186
Hispanic_00	3,402	1,971	1,763	0	11,200
Non-Hispanic White_00	3,402	1,879	1,596	0	12,173
Non-Hispanic Black_00	3,402	349	657	0	6,128
Non-Hispanic Indian_00	3,402	18	36	0	1,230
Non-Hispanic Asian_00	3,402	506	665	0	5,318
Non-Hispanic Other_00	3,402	132	119	0	1,073

Source: US Census Bureau, 1990 and 2000 Census Summary File 3 and 1990 CTPP.

Model Results

A multinomial logistic regression model of population by age group indicates that distribution probability of each age group in 2000 can be determined by the age distribution and other additional factors (Hispanic population, median household income, employment, population density) in 1990 (see table 3). The pseudo R-square of the estimated model is 0.283. The coefficients of most independent variables are significant for the estimation model for each age group. A multinomial logistic regression model of population by racial/ethnic group indicates that distribution probability of each

racial/ethnic group in 2000 can be determined by the racial/ethnic distribution and other additional factors (population of age25-64, median household income, employment, population density) in 1990 (see table 4). The pseudo R-square of the estimated model is 0.44, higher than the age model. The coefficients of most independent variables are significant for the estimation model for each age group.

Table 3. Results of a Multinomial Logistic Regression Model of Population by Age Group in 2000

Parameter	Age0_4		Age5_15		Age16		Age17		Age18_24		Age25_64	
	Coeff	S	Coeff	S	Coeff	S	Coeff	S	Coeff	S	Coeff	S
Intercept	-0.1216	***	0.6731	***	-1.9143	***	-1.8823	***	0.3568	***	1.5995	***
age0_4_90	0.0023	***	0.0018	***	0.0011	***	0.0011	***	0.0006	***	0.0008	***
age5_15_90	-0.0002	***	0.0003	***	0.0005	***	0.0004	***	0.0002	***	-0.0002	***
age16_90	-0.0006	***	-0.0001	***	0.0013	***	0.0012	***	0.0000	NS	-0.0006	***
age17_90	-0.0006	***	0.0000	NS	0.0011	***	0.0016	***	-0.0007	***	-0.0010	***
age18_24_90	0.0002	***	0.0001	***	0.0001	***	0.0002	***	0.0009	***	0.0003	***
age25_64_90	-0.0001	***	-0.0002	***	-0.0002	***	-0.0002	***	-0.0001	***	0.0001	***
age65_over_90	-0.0012	***	-0.0011	***	-0.0011	***	-0.0011	***	-0.0013	***	-0.0011	***
pop_his_90	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	**
median_ho_income_90	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	***
employment_90	0.0000	***	0.0000	***	0.0000	***	0.0000	NS	0.0000	***	0.0000	NS
density_p_90	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	***
-2 Log Likelihood Chi-S	628288.50											
DF	20000											
Pseudo R-Square	0.283											
Observations	3402											
Note: Coeff= Coefficient(β k: log-odds), S=Symbol, *** p<0.01, ** p<0.05, * p<0.01, NS= not significant												

Table 4. Results of a Multinomial Logistic Regression Model of Population by Race/Ethnicity in 2000

Parameter	Hispanic		NH White		NH Black		NH Indian		NH Asian	
	Coeff	S	Coeff	S	Coeff	S	Coeff	S	Coeff	S
Intercept	3.1999	***	2.6717	***	1.7284	***	-0.8733	***	0.8758	***
pop_his_90	0.0007	***	0.0000	***	0.0001	***	0.0003	***	0.0002	***
pop_white_nh_90	-0.0001	***	0.0002	***	-0.0002	***	0.0001	***	-0.0001	***
pop_black_nh_90	0.0000	***	-0.0005	***	0.0007	***	0.0001	***	-0.0004	***
pop_indian_nh_90	-0.0002	***	-0.0003	***	0.0000	NS	0.0037	***	-0.0025	***
pop_asian_nh_90	-0.0002	***	-0.0006	***	-0.0003	***	-0.0005	***	0.0010	***
pop_other_nh_90	-0.0004	***	-0.0006	***	0.0022	***	-0.0013	***	0.0001	NS
age25_64_90	-0.0003	***	-0.0002	***	0.0000	NS	-0.0002	***	-0.0002	***
median_ho_income_90	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	***
employment_90	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	***
density_p_90	0.0000	***	0.0000	***	0.0000	***	0.0000	***	0.0000	**
-2 Log Likelihood Chi-S	2291666.00									
DF	20000									
Pseudo R-Square	0.4407									
Observations	3402									
Note: Coeff = Coefficient(βk: log-odds), S=Symbol, *** p<0.01, ** p<0.05, * p<0.01, NS= not significant										

Model Application and Accuracy Measurement

The study uses a SCAG's draft regional growth forecast as a regional and county control for analysis purpose. A draft regional growth forecast was prepared for policy and planning analysis in February 2010. The draft regional growth forecast is a future snapshot of the most likely population and employment forecast at the regional level. It reflects historical trends, based on reasonable key technical assumptions. According to the draft regional growth forecast (see table 5), the region will add 4.4 million people to reach 23 million people by 2035. Supporting this population in 2035 will be a total of 9.7 million jobs in 2035 with 2 million new jobs. This level of population and job growth is expected to yield 1.5 million additional households in the region at an average of three persons per household.

Table 5. The Draft Regional Growth Forecast, 2008-2035. Unit: Thousands

	2008	2035	Change	% Change, 2008-2035
Population	18,626	23,005	4,379	24%
Households	5,866	7,346	1,480	25%
Employment	7,740	9,736	1,996	26%

Source: SCAG, Draft Growth Forecast, February 2010.

The regional population projection and its demographic characteristics will be the aging of population and shifts in ethnic distribution (see table 6). With the aging of the baby

boomer generation (born between 1946 and 1964), the share of the population 65 years old and over is projected to increase from 10 percent in 2008 to 16 percent in 2035, while the share of the population less than 65 years old decreases from 90 percent in 2010 to 84 percent in 2035. In particular, the share of the population of the working age 16–64 has its share sharply decline from 66.6 percent to 62.3 percent during the projection period. This implies a future shortage of workers. With the increasing share of the older population and the decreasing share of the working age population, the aged dependency ratio (i.e., the number of aged people per hundred people of working age) is projected to increase from 16 percent in 2008 to 26 percent in 2035 (an increase of 10 percent during the period). The other characteristic of the projected population is the racial/ethnic diversity. The region already has a high level of racial/ethnic diversity in 2008 with a Hispanic population of 45 percent, a non-Hispanic White population of 34 percent, a non-Hispanic Asian population and others of 14 percent, and a non-Hispanic Black population of 7 percent. The region’s racial/ethnic composition is projected to exhibit a rapid change toward a majority Hispanic population of 56 percent in 2035, while the share of the non-Hispanic White population is projected to drop sharply to 23 percent.

Table 6. Age and Racial/Ethnic Composition of Regional Population, 2008 & 2035

	2008		2035		2008-2035	
	Number	%	Number	%	Change	% Change
Age						
Age 0 4	1,363,983	7.3%	1,556,288	6.8%	192,305	14.1%
Age 5 15	2,943,071	15.8%	3,439,126	14.9%	496,055	16.9%
Age 16	291,480	1.6%	319,745	1.4%	28,265	9.7%
Age 17	291,482	1.6%	319,726	1.4%	28,244	9.7%
Age 18 24	2,039,116	10.9%	2,311,889	10.0%	272,773	13.4%
Age 25 64	9,771,433	52.5%	11,398,862	49.5%	1,627,429	16.7%
Age 65 over	1,925,525	10.3%	3,659,596	15.9%	1,734,071	90.1%
Race/Ethnicity						
Hispanic	8,362,683	44.9%	12,765,424	55.5%	4,402,741	52.6%
Non-Hispanic White	6,370,596	34.2%	5,365,363	23.3%	1,005,233	-15.8%
Non-Hispanic Black	1,292,655	6.9%	1,405,585	6.1%	112,930	8.7%
Non-Hispanic Indian	79,976	0.4%	115,496	0.5%	35,520	44.4%
Non-Hispanic Asian	2,068,008	11.1%	2,732,321	11.9%	664,313	32.1%
Non-Hispanic Other	452,172	2.4%	621,043	2.7%	168,871	37.3%
Total Population	18,626,090	100.0%	23,005,232	100.0%	4,379,142	23.5%

Source: SCAG, Draft Growth Forecast, February 2010.

Two growth scenarios are developed to allocate the regional growth into TAZs. The first scenario is a local input scenario, which represents the most likely growth and growth distribution of the region in the absence of the explicit regional policies. The existing local policies including zoning and general plan, and the locally supported Blueprint Planning land use policy are reflected in growth distribution. The most up-to-date local input forms the foundation of the local input scenario. The second scenario is a preferred plan scenario, which reflects land uses beyond what has been supported by local

jurisdictions. The TOD strategy is one example, and assigns greater capacity to areas around transit stations.

The local input scenario and the preferred plan scenario produce a different set of growth distributions, while maintaining the regional control. The TAZ distribution of 2035 population and households of two growth scenarios was analyzed using the mean absolute percentage error (MAPE) (see table 7). The mean absolute percentage error (MAPE) of the region-wide population and household distributions between two growth scenarios is 47% and 42%, respectively. There is a significant change in population and household distributions between two growth scenarios due to the region-wide application of the regional blueprint land use policy. Riverside County shows the largest discrepancy, while Orange County shows the smallest discrepancy. The data indicates that Orange County has already incorporated the Blueprint Planning land use policy in its local input scenario.

A similar approach was applied to age variables. 2035 TAZ population by age group was derived using the coefficients of the previous multi-nomial logit regression model. The total population of each age group was transformed into the relative percentage of each age group of population within TAZ for a fair comparison of two growth scenarios. The percent distribution of 2035 age groups of two growth scenarios at TAZ level was analyzed using the mean absolute percentage error (MAPE) (see table 7). The overall discrepancy of the region-wide age group distributions between two growth scenarios is low, ranging from 1.3% for age25-64 to 4.9% for age16. Riverside County generally shows the largest discrepancy, while Orange County shows the smallest discrepancy in the SCAG region.

As conducted in age variables analysis, the total population of each racial/ethnic group was transformed into the relative percentage of each racial/ethnic group of population within TAZ for a fair comparison of two growth scenarios. The percent distribution of 2035 racial/ethnic groups of two growth scenarios at TAZ level was analyzed using the mean absolute percentage error (MAPE) (see table 7). The overall discrepancy of the region-wide racial/ethnic group distributions between two growth scenarios is higher than that of age variables, ranging from 4.2% for Hispanic population to 19.5% for Non-Hispanic Indian population. Imperial County shows the largest discrepancy in the racial/ethnic categories, while Ventura County shows the smallest discrepancy in the SCAG region.

Table 7. MAPE for Model Results of Two Growth Scenarios for Year 2035

County	IMP	LA	OR	RIV	SB	VEN	SCAG Region
Observations	110	2,244	666	478	402	210	4,110
Population*	35.0%	20.1%	12.9%	250.8%	29.1%	15.8%	46.8%
Households*	40.3%	21.3%	8.6%	208.0%	28.8%	17.1%	42.0%
Age							
Age 0_4	9.1%	2.3%	2.6%	6.7%	4.1%	3.9%	3.3%
Age 5_15	4.5%	1.9%	1.6%	6.5%	3.0%	2.6%	2.6%
Age 16	8.9%	4.4%	2.8%	9.0%	5.6%	4.4%	4.9%
Age 17	8.8%	4.6%	2.9%	9.6%	5.7%	5.3%	5.1%
Age 18_24	5.5%	2.6%	2.2%	7.3%	3.6%	3.9%	3.3%
Age 25_64	1.5%	0.8%	0.5%	4.7%	1.8%	1.1%	1.3%
Age 65_ over	5.7%	2.8%	2.8%	11.7%	7.0%	5.3%	4.5%
Race/Ethnicity							
Hispanic	3.8%	3.5%	3.9%	7.1%	6.3%	3.2%	4.2%
Non-Hispanic White	11.4%	7.3%	5.1%	13.4%	12.1%	5.6%	8.1%
Non-Hispanic Black	22.0%	12.5%	14.9%	14.6%	12.6%	20.1%	13.8%
Non-Hispanic Indian	19.7%	18.4%	8.7%	18.8%	14.5%	12.5%	16.2%
Non-Hispanic Asian	28.0%	7.6%	10.6%	17.2%	13.4%	14.0%	10.6%
Non-Hispanic Other	16.4%	15.3%	36.8%	21.9%	13.2%	18.1%	19.5%

Note: * externally developed. IMP=Imperial, LA=Los Angeles, OR=Orange, RIV=Riverside, SB=San Bernardino, VEN=Ventura

The analysis of age distribution can be extended by using the dependency ratio, which requires only three different age groups for calculation. We use three age groups of 0-15, 16-64, 65+ for the analysis. There are three dependency ratios: the general dependency ratio (sum of age 0-15 and 65+ divided by age 16-64, then multiplied by 100), the youth dependency ratio (age 0-15 divided by age 16-64, then multiplied by 100), and the elderly dependency ratio (age 65+ divided by age 16-64, then multiplied by 100).

The percent difference of the average general, youth, and elderly dependency ratios for the county level (aggregated from TAZ dependency ratios) range from -0.1% to 5.5% with a region's figure of 1.0%, -1.0% to 3.6% with a region's figure of 0.3%, -0.3% to 7.2% with a region's figure of 1.7%, respectively. Overall the percent difference of region's average dependency ratios between two growth scenarios remains low, while the elderly dependency ratio shows much bigger variation than that of the youth dependency ratio. Riverside County shows the largest discrepancy in dependency ratios between growth scenarios, while Los Angeles and Orange Counties show the smallest discrepancy in the SCAG region. The significant shift of the small area population growth distributions between two growth scenarios might be related to the change in dependency ratios.

Table 7. Errors for Dependency Ratios of Two Growth Scenarios for Year 2035

County	TAZ	GDR			YDR			EDR		
	Number	L	P	%D	L	P	%D	L	P	%D
IMP	110	49.1	50.1	2.0%	26.3	26.4	0.3%	22.7	23.7	3.9%
LA	2,218	60.5	60.5	-0.1%	33.7	33.8	0.2%	26.8	26.7	-0.3%
OR	653	71.5	71.5	0.0%	33.9	33.6	-0.7%	37.7	37.9	0.7%
RIV	472	69.1	73.1	5.5%	32.9	34.1	3.6%	36.2	39.1	7.2%
SB	397	62.0	63.0	1.6%	36.7	36.4	-1.0%	25.3	26.6	5.1%
VEN	208	66.3	67.2	1.4%	32.7	32.9	0.8%	33.6	34.3	1.9%
SCAG Region	4,058	63.4	64.0	1.0%	33.7	33.8	0.3%	29.7	30.2	1.7%

Note: L= local input scenario, P= preferred plan scenario, %D = (P-L)/P*100,
GDR=general dependency ratio, YDR = youth dependency ratio, EDR=elderly dependency ratio

The analysis of racial/ethnic distribution can be conducted by using the entropy index, which measures the diversity of race/ethnic groups. We use four racial/ethnic groups of Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian & Others and Hispanic groups. The entropy index (unnormalized & normalized) can be computed using the following formula (Plane & Rogerson, 1994, p.302):

The unnormalized entropy index:

$$H = -\sum_{k=1}^n [(P_k / P) \ln(P_k / P)]$$

Where

n =number of racial/ethnic groups,

P_k = population of the k th racial/ethnic group,

P = total population

The normalized entropy index shows the range of all values from 0 to 1. The formula is

$$H^* = H / \ln n$$

The percent difference of the normalized Entropy Index for the county level (aggregated from TAZ Entropy Index) range from -0.5% to 1.2% with a region's figure of -0.5%. Overall the percent difference of region's normalized Entropy Index between two growth scenarios remains low. Riverside County shows the largest discrepancy in the normalized Entropy Index between growth scenarios, while San Bernardino and Ventura Counties show the smallest discrepancy in the SCAG region. The significant shift of the small area population growth distributions between two growth scenarios might be related to the change in Entropy Index.

Table 8. Errors for Entropy Index of Two Growth Scenarios for Year 2035x

County	TAZ	Entropy Index (unnormalized)			Entropy Index (normalized)		
	Number	L	P	%D	L	P	%D
IMP	103	0.708	0.705	-0.5%	0.511	0.508	-0.5%
LA	2,196	0.833	0.824	-1.1%	0.601	0.594	-1.1%
OR	638	0.912	0.905	-0.7%	0.658	0.653	-0.7%
RIV	470	0.989	1.001	1.2%	0.713	0.722	1.2%
SB	392	1.023	1.024	0.1%	0.738	0.739	0.1%
VEN	197	0.813	0.814	0.1%	0.586	0.587	0.1%
SCAG Region	3,996	0.878	0.874	-0.5%	0.633	0.630	-0.5%

L= local input scenario, P= preferred plan scenario, %D = (P-L)/P*100

In summary, individual age variables maintain a relatively low variation in the difference of the age distribution between two growth scenarios. They tend to show a noticeable gap in the difference of the age distribution between two growth scenarios among six counties in the SCAG region. Individual race/ethnicity variables tend to show a relatively high variation in the difference of the racial/ethnic distribution between two growth scenarios. They also show a noticeable gap in the difference of the racial/ethnic distribution between two growth scenarios among six counties in the SCAG region. When presented in demographic indicators, including dependency ratios and entropy index, the error pattern changed. The racial/ethnic distributions tend to show more difference between two growth scenarios than the age distributions. It should be noted that the more refined analysis might be needed due to availability of many TAZs with small population and household figures. The presence of many such TAZs might skew the summary error statistics.

Conclusions

The small area population projections are important in understanding the diverse community service needs of the future. Although the population size of the small area might be a useful indicator for measuring the community needs, the detailed demographic projections, if available, would be able to more accurately estimate the community service needs. This study presents a coherent modeling framework and a statistical approach for projecting the population size and demographic characteristics of the small area within the metropolitan planning context.

First, the popular cohort-component method is not easily applicable to population projections of small area (e.g., census tract or transportation analysis zone) for a couple of reasons: (1) the historical and current trends in vital statistics and migration of the small areas are not easily available; (2) it is difficult to independently develop reasonable migration assumptions of small areas. However, housing development and assumptions are easily available from the building permit pattern and land use assumptions from the local general plan or the alternative scenarios. A reasonable approach is to derive population projections of local communities through the enhanced linkage of small area housing growth and regional demographic processes. If this framework is accepted, then the small area population and demographic projections can be derived using both the

demographic process at the large area level and the availability of housing at the small area level. The small area demographic projections can be a useful baseline framework for doing diverse applied demographic and planning research in the following areas: aging in place; gentrification; identification of ethnic residential enclave; environmental justice analysis; quantification of planning efforts on small area demography, etc.

Second, this study proposes a statistical multi-nomial logit regression method to develop more detailed demographic characteristics of projected population at the TAZ level. The presented regression approach showed several advantages in terms of maintaining a consistency with the large area's demographic pattern, and the small area population size, and also the historical pattern of the demographic characteristics of the large area. Through the comparison analysis of mean absolute percentage error of age and racial/ethnic distribution for two different growth scenarios, the presented regression method is found to produce the low variation in the age distribution and the moderate variation in the racial/ethnic distribution. Overall the study finds the proposed method as reasonable, and suggests that the approach might need to be further enhanced to minimize the variation in the age and racial/ethnic distribution

Finally, the presented approach basically assumes that the community tends to maintain the community's existing nature during the projection period. Therefore, the presented regress approach may not properly reflect the newly emerging demographic attributes of projected local population as a result of urban development. The presented approach may not model the implication of urban infill development: gentrification, and the related demographic change from low-middle income to middle-high income. Probably the specially designed modeling approach might be needed to identify the emerging demographic changes of projected population associated with the specific development activity (e.g., TOD).

Acknowledgment

Author wishes to acknowledge SungHo Ryu for preparing tables 1 and 2 used in this paper.

References

- Brail, Richard K., R.E. Klosterman (eds.). 2001. Planning Support Systems. Redlands, California: ESRI.
- Eluru, Naveen, A. Pinjari, J. Guo, I. Sener, S. Srinivasan, R. Copperman, C. Bhat. 2008. Population Updating System Structures and Models Embedded in the Comprehensive Econometric Microsimulator for Urban Systems. *Transportation Research Record* 2076: 171-182.
- Field, Brian., B. MacGregor, 1987. *Forecasting Techniques for Urban and Regional Planning*. London: UCL Press
- Hopkins, Lewis D., M.A. Zapata. (eds.). 2007. *Engaging the Future – Forecasts, Scenarios, Plans, and Projects*. Cambridge, Massachusetts: Lincoln Institute of Land Policy.
- Kanaroglou, P.S., H.F.Maoh, B. Newbold, D.M. Scott, A. Paez. 2009. A Demographic Model for Small Area Population Projections: an Application to the Census Metropolitan Area of Hamilton in Ontario, Canada. *Environment and Planning A*, Volume 41, pp. 964-979.
- Klosterman, Richard E. 1990. *Community and Analysis Planning Techniques*. Rowmand and Littlefield Publishers, Inc. Savage, Maryland. See chapter 4-8.
- Liao, Tim Futing. 1994. *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*. (Sage University Paper Series on Quantitative Applications in the Social Sciences, Series no. 07-101). Thousand Oaks, CA: Sage.
- Lowry, Ira S. 1964. *A Model of Metropolis* RAND Memorandum 4025-RC.
- Menard, Scott. 2002. *Applied Logistic Regression Analysis*. Second Edition. (Sage University Paper Series on Quantitative Applications in the Social Sciences, Series no. 07-106). Thousand Oaks, CA: Sage.
- Park, Heonsoo. 2009. *Advanced Programming Support for Developing Traffic Analysis Zone (TAZ)/Grid Cell Socio-economic Data and Assessing Selected Small Area Allocation Models for Southern California Association of Governments (unpublished)*.
- Pampel, Fred. C. 2000. *Logistic Regression: A Primer*. (Sage University Paper Series on Quantitative Applications in the Social Sciences, Series no. 07-106). Thousand Oaks, CA: Sage.

- Pitkin, John and D. Myers. 2011. "A Period Summary Measure of Immigrant Advancement," *Demographic Research* 24: 257-92
<http://www.demographic-research.org/volumes/vol24/12/>.
- Pagliara, Francesca, J. Preston, and D. Simmonds (eds). 2010) *Residential Location Choice: Models and Applications*. Heidelberg: Springer.
- Plane, David., P.A. Rogerson. 1994. *The Geographical Analysis of Population: With Applications to Planning and Business*. New York: John Wiley & Sons
- Putman, Stephen H. 2010. DRAM Residential Location and Land Use Model: 40 Years of Development and Application, in *Residential Location Choice: Models and Applications*, Pagliara, Francesca, J. Preston, and D. Simmonds (eds). Heidelberg: Springer.
- Rees, P., P. Norman, D. Brown. 2004. A framework for progressively improving small area population estimates, *Journal of Royal Statistical Society A* 167(1): 5-36
- Simpson, L., M. Tranmer. 2005. Combining sample and census data in small area estimates: iterative proportional fitting with standard software, *The Professional Geographer* 57: 222- 234.
- Smith, Stanley K., B.B. Lewis. 1983. Some New Techniques for Applying the Housing Unit Method of Local Population Estimation: Further Evidence, *Demography* Vol. 20 No. 3: 407-413.
[http://www.bebr.ufl.edu/files/1983%20Demog%20\(Further%20Evidence\).pdf](http://www.bebr.ufl.edu/files/1983%20Demog%20(Further%20Evidence).pdf)
- Smith, Stanley K., J. Tayman, D. A. Swanson. 2001. *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.
- Southern California Association of Governments (SCAG). 2008. 2008 Regional Transportation Plan: Making the Connection, Growth Forecast Report.
<http://www.scag.ca.gov/rtp2008/pdfs/finalrtp/reports/fGrowthForecast.pdf>
- Southern California Association of Governments (SCAG). 2010. SCAG Draft Growth Forecast.